

Numériser le patrimoine I: standards et bonnes pratiques

L'analyse de mise en page

Simon Gabay

Genève



Pre-processing

Rotation

Original	Résultat
<p style="text-align: center;">— 6 —</p> <p>Il lui fait part du décès de Gaston d'Orléans; détails sur la fin du prince. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87545 BIGOTTINI (Émilie), célèbre danseuse, qui eut de grands succès à l'Opéra. — L. a. s. à M^{me} Guyet; 29 août 1833, 10 " 2 p. in-folio. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87546 BLANQUI (Jérôme-Adolphe), économiste, membre de l'Académie des Sciences morales et politiques, né à Nice. — 30 l. a. s. à Jullien de Paris; 1818-1841, 50 p. env. in-4° ou in-8°. 30 " Intéressante correspondance toute relative à ses travaux et à sa collaboration à la <i>Revue encyclopédique</i>. Il parle de ses frères dont il dirigeait l'éducation, des moyens d'échapper à la conscription. « Mes parents sont dans l'impossibilité de m'acheter un remplaçant, et il me serait bien désagréable d'interrompre ma carrière et de perdre le fruit de mes études entrant comme simple soldat dans les rangs de l'armée. » Il demande la réintégration dans le conventionnel, un des 73 du 3^e mai dont il retrace une brève biographie. Il accepte de collaborer à la <i>Revue encyclopédique</i>, mais il refuse de s'occuper de la rédaction des tables; il demande des renseignements pour écrire une vie de Kosozuko, l'émule de Washington. Les lettres postérieures traitent plus particulièrement d'économie sociale.</p> <p>87547 BONAPARTE (Jérôme), le fils du roi Jérôme. — L. a. s. N. à M. Gasc; Paris, 31 octobre 1848, 2 p. 1/2 in-8°. 35 " Curieuse lettre politique entièrement relative à la préparation de l'élection de son cousin à la présidence de la République. « Tous les chefs sont pour nous, nos vrais appuis sont les paysans, les ouvriers et les soldats, la tête de tous les partis nous est hostile, cependant les hommes d'ordres reviennent à nous. Cavaignac a pour lui le pouvoir, la corruption et l'intimidation; son cousin qui voit (sic) que lui, mieux que tout autre peut sauver la République, l'établir à tout jamais en France! »</p> <p>87548 BOSQUET (Pierre-Jean-François), maréchal de France, qui s'illustra à l'Alma, à Inkermann, et à l'assaut de la tour Malakoff, né à Mont-de-Marsan (Landes). — L. a. s. à M. Renoux; 12 juin 1851, 3 p. in-8°. 25 " Intéressante lettre entièrement relative aux opérations militaires; il se plaint de la froideur de ses vieux camarades, qui jaloussent ses succès.</p> <p>87549 BRUANT (Aristide), chansonnier montmartrois, né à Courtenay (Loiret). — L. a. s., à un confrère; Le Pouldu, 15 août 1911, 1 p. in-8°. 5 " Il dit que sa pièce les <i>Bas-fonds de Paris</i> est déjà faite; il l'a écrite en collaboration avec Arthur Bernède.</p> <p>87550 BRUEYS (François-Paul), vice-amiral, commandant en</p>	<p style="text-align: center;">— 6 —</p> <p>Il lui fait part du décès de Gaston d'Orléans; détails sur la fin du prince. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87545 BIGOTTINI (Émilie), célèbre danseuse, qui eut de grands succès à l'Opéra. — L. a. s. à M^{me} Guyet; 29 août 1833, 10 " 2 p. in-folio. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87546 BLANQUI (Jérôme-Adolphe), économiste, membre de l'Académie des Sciences morales et politiques, né à Nice. — 30 l. a. s. à Jullien de Paris; 1818-1841, 50 p. env. in-4° ou in-8°. 30 " Intéressante correspondance toute relative à ses travaux et à sa collaboration à la <i>Revue encyclopédique</i>. Il parle de ses frères dont il dirigeait l'éducation, des moyens d'échapper à la conscription. « Mes parents sont dans l'impossibilité de m'acheter un remplaçant, et il me serait bien désagréable d'interrompre ma carrière et de perdre le fruit de mes études entrant comme simple soldat dans les rangs de l'armée. » Il demande la réintégration dans le conventionnel, un des 73 du 3^e mai dont il retrace une brève biographie. Il accepte de collaborer à la <i>Revue encyclopédique</i>, mais il refuse de s'occuper de la rédaction des tables; il demande des renseignements pour écrire une vie de Kosozuko, l'émule de Washington. Les lettres postérieures traitent plus particulièrement d'économie sociale.</p> <p>87547 BONAPARTE (Jérôme), le fils du roi Jérôme. — L. a. s. N. à M. Gasc; Paris, 31 octobre 1848, 2 p. 1/2 in-8°. 35 " Curieuse lettre politique entièrement relative à la préparation de l'élection de son cousin à la présidence de la République. « Tous les chefs sont pour nous, nos vrais appuis sont les paysans, les ouvriers et les soldats, la tête de tous les partis nous est hostile, cependant les hommes d'ordres reviennent à nous. Cavaignac a pour lui le pouvoir, la corruption et l'intimidation; son cousin qui voit (sic) que lui, mieux que tout autre peut sauver la République, l'établir à tout jamais en France! »</p> <p>87548 BOSQUET (Pierre-Jean-François), maréchal de France, qui s'illustra à l'Alma, à Inkermann, et à l'assaut de la tour Malakoff, né à Mont-de-Marsan (Landes). — L. a. s. à M. Renoux; 12 juin 1851, 3 p. in-8°. 25 " Intéressante lettre entièrement relative aux opérations militaires; il se plaint de la froideur de ses vieux camarades, qui jaloussent ses succès.</p> <p>87549 BRUANT (Aristide), chansonnier montmartrois, né à Courtenay (Loiret). — L. a. s., à un confrère; Le Pouldu, 15 août 1911, 1 p. in-8°. 5 " Il dit que sa pièce les <i>Bas-fonds de Paris</i> est déjà faite; il l'a écrite en collaboration avec Arthur Bernède.</p> <p>87550 BRUEYS (François-Paul), vice-amiral, commandant en</p>

Niveau de gris

Original	Résultat
<p style="text-align: center;">— 6 —</p> <p>Il lui fait part du décès de Gaston d'Orléans; détails sur la fin du prince. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87545 BIGOTTINI (Émilie), célèbre danseuse, qui eut de grands succès à l'Opéra. — L. a. s. à M^{me} Guyet; 29 aout 1833. 2 p. in-folio. 10 °</p> <p>Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87546 BLANQUI (Jérôme-Adolphe), économiste, membre de l'Académie des Sciences morales et politiques, né à Nice. — 30 l. a. s. à Jullien de Paris; 1818-1841. 50 p. env. in-4° ou in-8°. 30 °</p> <p>Intéressante correspondance toute relative à ses travaux et à sa collaboration à la <i>Revue encyclopédique</i>... Il parle de ses frères dont il dirigeait l'éducation, des moyens d'échapper à la conscription. « Mes parents sont dans l'impossibilité de m'acheter un remplaçant, et il me serait bien désagréable d'interrompre ma carrière et de perdre le fruit de mes études entrant comme simple soldat dans les rangs de l'armée. » Il demande la permission de faire partie de l'armée le 29 ou 31 mai, dont il reçoit une brève biographie. Il accepte de collaborer à la <i>Revue encyclopédique</i>, mais il refuse de s'occuper de la rédaction des tables; il demande des renseignements pour écrire une vie de Kosciusko, l'émule de Washington. Les lettres postérieures traitent plus particulièrement d'économie sociale.</p> <p>87547 BONAPARTE (Jérôme), le fils du roi Jérôme. — L. a. s. N. à M. Gasc; Paris, 21 octobre 1848. 2 p. 1/2 in-8°. 35 °</p> <p>Curieuse lettre politique entièrement relative à la préparation de l'élection de son cousin à la présidence de la République. « Tous les chefs sont mal pour nous, nos vrais appuis sont les paysans, les ouvriers et les soldats, la tête de tous les partis nous est hostile, cependant les hommes d'ordre reviennent à nous. Cavaignac a pour lui le pouvoir, la corruption et l'intimidation; son cousin a pour lui le nom de l'Empereur et la sympathie des masses » qui voit (sic) que lui, mieux que tout autre peut sauver la République, établir à tout jamais en France !</p> <p>87548 BOSQUET (Pierre-Jean-François), maréchal de France, qui s'illustre à l'Alma, à Inkermann, et à l'assaut de la tour Malakoff, né à Mont-de-Marsan (Landes). — L. a. s. à M. Renoux; 12 juin 1851. 3 p. in-8°. 25 °</p> <p>Intéressante lettre entièrement relative aux opérations militaires; il se plaint de la froideur de ses vieux camarades, qui jaloussent ses succès.</p> <p>87549 BRUANT (Aristide), chansonnier montmartrois, né à Courtenay (Loiret). — L. a. s., à un confrère; Le Pouldu, 15 aout 1911. 1 p. in-8°. 5 °</p> <p>Il dit que sa pièce les <i>Bas-fonds de Paris</i> est déjà faite; il l'a écrite en collaboration avec Arthur Bernède.</p> <p>87550 BRUEYS (François-Paul), vice-amiral, commandant en</p>	<p style="text-align: center;">— 6 —</p> <p>Il lui fait part du décès de Gaston d'Orléans; détails sur la fin du prince. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87545 BIGOTTINI (Émilie), célèbre danseuse, qui eut de grands succès à l'Opéra. — L. a. s. à M^{me} Guyet; 29 aout 1833. 2 p. in-folio. 10 °</p> <p>Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87546 BLANQUI (Jérôme-Adolphe), économiste, membre de l'Académie des Sciences morales et politiques, né à Nice. — 30 l. a. s. à Jullien de Paris; 1818-1841. 50 p. env. in-4° ou in-8°. 30 °</p> <p>Intéressante correspondance toute relative à ses travaux et à sa collaboration à la <i>Revue encyclopédique</i>... Il parle de ses frères dont il dirigeait l'éducation, des moyens d'échapper à la conscription. « Mes parents sont dans l'impossibilité de m'acheter un remplaçant, et il me serait bien désagréable d'interrompre ma carrière et de perdre le fruit de mes études entrant comme simple soldat dans les rangs de l'armée. » Il demande la permission de faire partie de l'armée le 29 ou 31 mai, dont il reçoit une brève biographie. Il accepte de collaborer à la <i>Revue encyclopédique</i>, mais il refuse de s'occuper de la rédaction des tables; il demande des renseignements pour écrire une vie de Kosciusko, l'émule de Washington. Les lettres postérieures traitent plus particulièrement d'économie sociale.</p> <p>87547 BONAPARTE (Jérôme), le fils du roi Jérôme. — L. a. s. N. à M. Gasc; Paris, 21 octobre 1848. 2 p. 1/2 in-8°. 35 °</p> <p>Curieuse lettre politique entièrement relative à la préparation de l'élection de son cousin à la présidence de la République. « Tous les chefs sont mal pour nous, nos vrais appuis sont les paysans, les ouvriers et les soldats, la tête de tous les partis nous est hostile, cependant les hommes d'ordre reviennent à nous. Cavaignac a pour lui le pouvoir, la corruption et l'intimidation; son cousin a pour lui le nom de l'Empereur et la sympathie des masses » qui voit (sic) que lui, mieux que tout autre peut sauver la République, établir à tout jamais en France !</p> <p>87548 BOSQUET (Pierre-Jean-François), maréchal de France, qui s'illustre à l'Alma, à Inkermann, et à l'assaut de la tour Malakoff, né à Mont-de-Marsan (Landes). — L. a. s. à M. Renoux; 12 juin 1851. 3 p. in-8°. 25 °</p> <p>Intéressante lettre entièrement relative aux opérations militaires; il se plaint de la froideur de ses vieux camarades, qui jaloussent ses succès.</p> <p>87549 BRUANT (Aristide), chansonnier montmartrois, né à Courtenay (Loiret). — L. a. s., à un confrère; Le Pouldu, 15 aout 1911. 1 p. in-8°. 5 °</p> <p>Il dit que sa pièce les <i>Bas-fonds de Paris</i> est déjà faite; il l'a écrite en collaboration avec Arthur Bernède.</p> <p>87550 BRUEYS (François-Paul), vice-amiral, commandant en</p>

Binarisation

Original	Résultat
<p style="text-align: center;">— 6 —</p> <p>Il lui fait part du décès de Gaston d'Orléans; détails sur la fin du prince. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87545 BIGOTTINI (Émilie), célèbre danseuse, qui eut de grands succès à l'Opéra. — L. a. s. à M^{me} Guyet; 29 aout 1833, 2 p. in-folio. — 10 " 2 p. in-folio. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87546 BLANQUI (Jérôme-Adolphe), économiste, membre de l'Académie des Sciences morales et politiques, né à Nice; 30 l. a. s. à Jullien de Paris; 1818-1841, 50 p. env. in-4° ou in-8°. 30 " Intéressante correspondance toute relative à ses travaux et à sa collaboration à la <i>Revue encyclopédique</i>... Il parle de ses frères dont il dirigeait l'éducation, des moyens d'échapper à la conscription. « Mes parents sont dans l'impossibilité de m'acheter un remplaçant, et il me serait bien désagréable d'interrompre ma carrière et de perdre le fruit de mes études entrant comme simple soldat dans les rangs de l'armée. » Il demande la permission de faire partie de l'armée le 22 du 31 mai, dont il reçoit une brève biographie. Il accepte de collaborer à la <i>Revue encyclopédique</i>, mais il refuse de s'occuper de la rédaction des tables; il demande des renseignements pour écrire une vie de Kosciusko, l'émule de Washington. Les lettres postérieures traitent plus particulièrement d'économie sociale.</p> <p>87547 BONAPARTE (Jérôme), le fils du roi Jérôme. — L. a. s. N. à M. Gasc; Paris, 31 octobre 1848, 2 p. 1/2 in-8°. 35 " Curieuse lettre politique entièrement relative à la préparation de l'élection de son cousin à la présidence de la République. « Tous les chefs sont mal pour nous, nos vrais appuis sont les paysans, les ouvriers et les soldats, la tête de tous les partis nous est hostile, cependant les hommes d'ordre reviennent à nous. Cavaignac a pour lui le pouvoir, la corruption et l'intimidation; son cousin a pour lui le nom de l'Empereur et la sympathie des masses » qui voit (sic) que lui, mieux que tout autre peut sauver la République, établir à tout jamais en France ! »</p> <p>87548 BOSQUET (Pierre-Jean-François), maréchal de France, qui s'illustra à l'Alma, à Inkermann, et à l'assaut de la tour Malakoff, né à Mont-de-Marsan (Landes). — L. a. s. à M. Renoux; 12 juin 1851, 3 p. in-8°. 25 " Intéressante lettre entièrement relative aux opérations militaires; il se plaint de la froideur de ses vieux camarades, qui jaloussent ses succès.</p> <p>87549 BRUANT (Aristide), chansonnier montmartrois, né à Courtenay (Loiret). — L. a. s., à un confrère; Le Pouldu, 15 aout 1911, 1 p. in-8°. 5 " Il dit que sa pièce les <i>Bas-fonds de Paris</i> est déjà faite; il l'a écrite en collaboration avec Arthur Bernède.</p> <p>87550 BRUEYS (François-Paul), vice-amiral, commandant en</p> <p style="text-align: center;">— 6 —</p> <p>Il lui fait part du décès de Gaston d'Orléans; détails sur la fin du prince. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87545 BIGOTTINI (Émilie), célèbre danseuse, qui eut de grands succès à l'Opéra. — L. a. s. à M^{me} Guyet; 29 aout 1833, 2 p. in-folio. — 10 " 2 p. in-folio. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87546 BLANQUI (Jérôme-Adolphe), économiste, membre de l'Académie des Sciences morales et politiques, né à Nice; 30 l. a. s. à Jullien de Paris; 1818-1841, 50 p. env. in-4° ou in-8°. 30 " Intéressante correspondance toute relative à ses travaux et à sa collaboration à la <i>Revue encyclopédique</i>... Il parle de ses frères dont il dirigeait l'éducation, des moyens d'échapper à la conscription. « Mes parents sont dans l'impossibilité de m'acheter un remplaçant, et il me serait bien désagréable d'interrompre ma carrière et de perdre le fruit de mes études entrant comme simple soldat dans les rangs de l'armée. » Il demande la permission de faire partie de l'armée le 22 du 31 mai, dont il reçoit une brève biographie. Il accepte de collaborer à la <i>Revue encyclopédique</i>, mais il refuse de s'occuper de la rédaction des tables; il demande des renseignements pour écrire une vie de Kosciusko, l'émule de Washington. Les lettres postérieures traitent plus particulièrement d'économie sociale.</p> <p>87547 BONAPARTE (Jérôme), le fils du roi Jérôme. — L. a. s. N. à M. Gasc; Paris, 31 octobre 1848, 2 p. 1/2 in-8°. 35 " Curieuse lettre politique entièrement relative à la préparation de l'élection de son cousin à la présidence de la République. « Tous les chefs sont mal pour nous, nos vrais appuis sont les paysans, les ouvriers et les soldats, la tête de tous les partis nous est hostile, cependant les hommes d'ordre reviennent à nous. Cavaignac a pour lui le pouvoir, la corruption et l'intimidation; son cousin a pour lui le nom de l'Empereur et la sympathie des masses » qui voit (sic) que lui, mieux que tout autre peut sauver la République, établir à tout jamais en France ! »</p> <p>87548 BOSQUET (Pierre-Jean-François), maréchal de France, qui s'illustra à l'Alma, à Inkermann, et à l'assaut de la tour Malakoff, né à Mont-de-Marsan (Landes). — L. a. s. à M. Renoux; 12 juin 1851, 3 p. in-8°. 25 " Intéressante lettre entièrement relative aux opérations militaires; il se plaint de la froideur de ses vieux camarades, qui jaloussent ses succès.</p> <p>87549 BRUANT (Aristide), chansonnier montmartrois, né à Courtenay (Loiret). — L. a. s., à un confrère; Le Pouldu, 15 aout 1911, 1 p. in-8°. 5 " Il dit que sa pièce les <i>Bas-fonds de Paris</i> est déjà faite; il l'a écrite en collaboration avec Arthur Bernède.</p> <p>87550 BRUEYS (François-Paul), vice-amiral, commandant en</p>	

Binarisations

Le principe de la binarisation d'une image est de déterminer un seuil: les pixels dont le niveau de gris est en dessous du seuil deviennent noirs, et ceux au dessus deviennent blanc.

On n'utilise plus de seuil global (i.e. pour la page) mais par exemple des seuils contextuels calculés à partir de la moyenne et de la cote z d'une fenêtre centrée sur le pixel. Il existe une multitude de techniques plus ou moins efficaces comme Niblack ou son amélioration par Sauvola (cf. image).

Niblack Threshold

Region-based segmentation

Let us first determine markers of the coins and the background. These markers are pixels that we can label unambiguously as either object or background. Here, the markers are found at the two extreme parts of the histogram of grey values:

```
>>> markers = np.zeros_like(coins)
```

Sauvola Threshold

Region-based segmentation

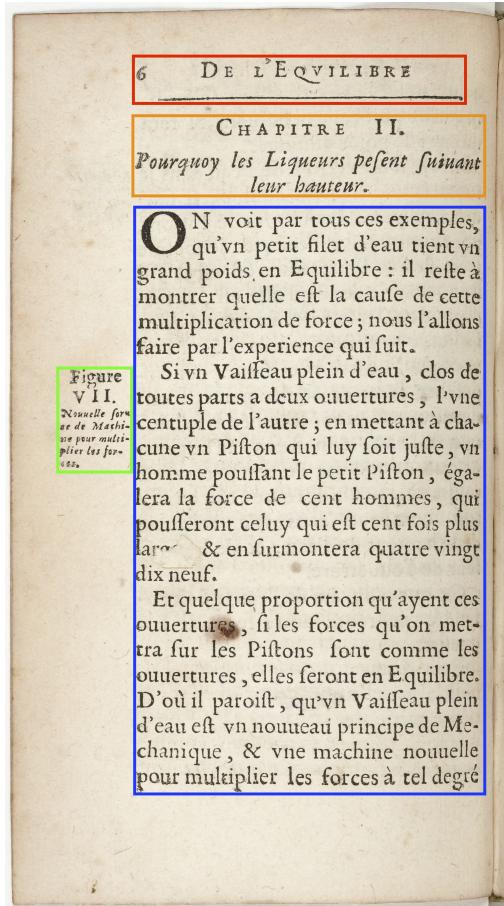
Let us first determine markers of the coins and the background. These markers are pixels that we can label unambiguously as either object or background. Here, the markers are found at the two extreme parts of the histogram of grey values:

```
>>> markers = np.zeros_like(coins)
```

Segmentation I

Original	Résultat
<p style="text-align: center;">— 6 —</p> <p>Il lui fait part du décès de Gaston d'Orléans; détails sur la fin du prince. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87545 BIGOTTINI (Émilie), célèbre danseuse, qui eut de grands succès à l'Opéra. — L. a. s. à M^{me} Guyet; 29 août 1833, 2 p. in-folio. 10 " Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87546 BLANOUI (Jérôme-Adolphe), économiste, membre de l'Académie des Sciences morales et politiques, né à Nice; 30 l. a. s. à Jullien de Paris; 1818-1841, 50 p. env., 1^{re} éd. in-8°. 30 " Intéressante correspondance toute relative à ses travaux et à sa collaboration à la <i>Revue encyclopédique</i>... Il parle de ses frères dont il dirigeait l'éducation, des moyens d'échapper à la conscription. « Mes parents sont dans l'impossibilité de m'acheter un remplaçant, et il me serait bien désagréable d'interrompre ma carrière et de perdre le fruit de mes études en entrant comme simple soldat dans les rangs de l'armée... » Il demande la rémission de son solde au 23 juillet 1832, 22 de 21 mai, dont il retrace une brève biographie. Il accepte de collaborer à la <i>Revue encyclopédique</i>, mais il refuse de s'occuper de la rédaction des tables; il demande des renseignements pour écrire une vie de Kosciusko, l'émule de Washington. Les lettres postérieures traitent plus particulièrement d'économie sociale.</p> <p>87547 BONAPARTE (Jérôme), le fils du roi Jérôme. — L. a. s. à M. Gasc; Paris, 31 octobre 1848, 2 p. 1/2 in-8°. 35 " Curieuse lettre politique entièrement relative à la préparation de l'élection de son cousin à la présidence de la République. « Tous les chefs sont mal pour nous, nos vrais appuis sont les paysans, les ouvriers et les soldats, la tête de tous les partis nous est hostile, cependant les hommes d'ordre reviennent à nous. Cavaignac a pour lui le pouvoir, la corruption et l'intimidation; son cousin a pour lui le nom de l'Empereur et la sympathie des masses » qui voit (sic) que lui, mieux que tout autre peut sauver la République, établir à tout jamais en France !</p> <p>87548 BOSQUET (Pierre-Jean-François), maréchal de France, qui s'illustre à l'Alma, à Inkermann, et à l'assaut de la tour Malakoff, né à Mont-de-Marsan (Landes). — L. a. s. à M. Renoux; 12 juin 1851, 3 p. in-8°. 25 " Intéressante lettre entièrement relative aux opérations militaires; il se plaint de la froideur de ses vieux camarades, qui jaloussent ses succès.</p> <p>87549 BRUANT (Aristide), chansonnier montmartrois, né à Courtenay (Loiret). — L. a. s., à un confrère; Le Pouldu, 15 août 1911, 1 p. in-8°. 5 " Il dit que sa pièce les <i>Bas-fonds de Paris</i> est déjà faite; il l'a écrite en collaboration avec Arthur Bernède.</p> <p>87550 BRUEYS (François-Paul), vice-amiral, commandant en</p>	<p style="text-align: center;">— 6 —</p> <p>Il lui fait part du décès de Gaston d'Orléans; détails sur la fin du prince. Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87545 BIGOTTINI (Émilie), célèbre danseuse, qui eut de grands succès à l'Opéra. — L. a. s. à M^{me} Guyet; 29 août 1833, 2 p. in-folio. 10 " Chaleureuse recommandation en faveur d'une femme de chambre.</p> <p>87546 BLANOUI (Jérôme-Adolphe), économiste, membre de l'Académie des Sciences morales et politiques, né à Nice; 30 l. a. s. à Jullien de Paris; 1818-1841, 50 p. env., 1^{re} éd. in-8°. 30 " Intéressante correspondance toute relative à ses travaux et à sa collaboration à la <i>Revue encyclopédique</i>... Il parle de ses frères dont il dirigeait l'éducation, des moyens d'échapper à la conscription. « Mes parents sont dans l'impossibilité de m'acheter un remplaçant, et il me serait bien désagréable d'interrompre ma carrière et de perdre le fruit de mes études en entrant comme simple soldat dans les rangs de l'armée... » Il demande la rémission de son solde au 23 juillet 1832, 22 de 21 mai, dont il retrace une brève biographie. Il accepte de collaborer à la <i>Revue encyclopédique</i>, mais il refuse de s'occuper de la rédaction des tables; il demande des renseignements pour écrire une vie de Kosciusko, l'émule de Washington. Les lettres postérieures traitent plus particulièrement d'économie sociale.</p> <p>87547 BONAPARTE (Jérôme), le fils du roi Jérôme. — L. a. s. à M. Gasc; Paris, 31 octobre 1848, 2 p. 1/2 in-8°. 35 " Curieuse lettre politique entièrement relative à la préparation de l'élection de son cousin à la présidence de la République. « Tous les chefs sont mal pour nous, nos vrais appuis sont les paysans, les ouvriers et les soldats, la tête de tous les partis nous est hostile, cependant les hommes d'ordre reviennent à nous. Cavaignac a pour lui le pouvoir, la corruption et l'intimidation; son cousin a pour lui le nom de l'Empereur et la sympathie des masses » qui voit (sic) que lui, mieux que tout autre peut sauver la République, établir à tout jamais en France !</p> <p>87548 BOSQUET (Pierre-Jean-François), maréchal de France, qui s'illustre à l'Alma, à Inkermann, et à l'assaut de la tour Malakoff, né à Mont-de-Marsan (Landes). — L. a. s. à M. Renoux; 12 juin 1851, 3 p. in-8°. 25 " Intéressante lettre entièrement relative aux opérations militaires; il se plaint de la froideur de ses vieux camarades, qui jaloussent ses succès.</p> <p>87549 BRUANT (Aristide), chansonnier montmartrois, né à Courtenay (Loiret). — L. a. s., à un confrère; Le Pouldu, 15 août 1911, 1 p. in-8°. 5 " Il dit que sa pièce les <i>Bas-fonds de Paris</i> est déjà faite; il l'a écrite en collaboration avec Arthur Bernède.</p> <p>87550 BRUEYS (François-Paul), vice-amiral, commandant en</p>

Segmentation: mise en page

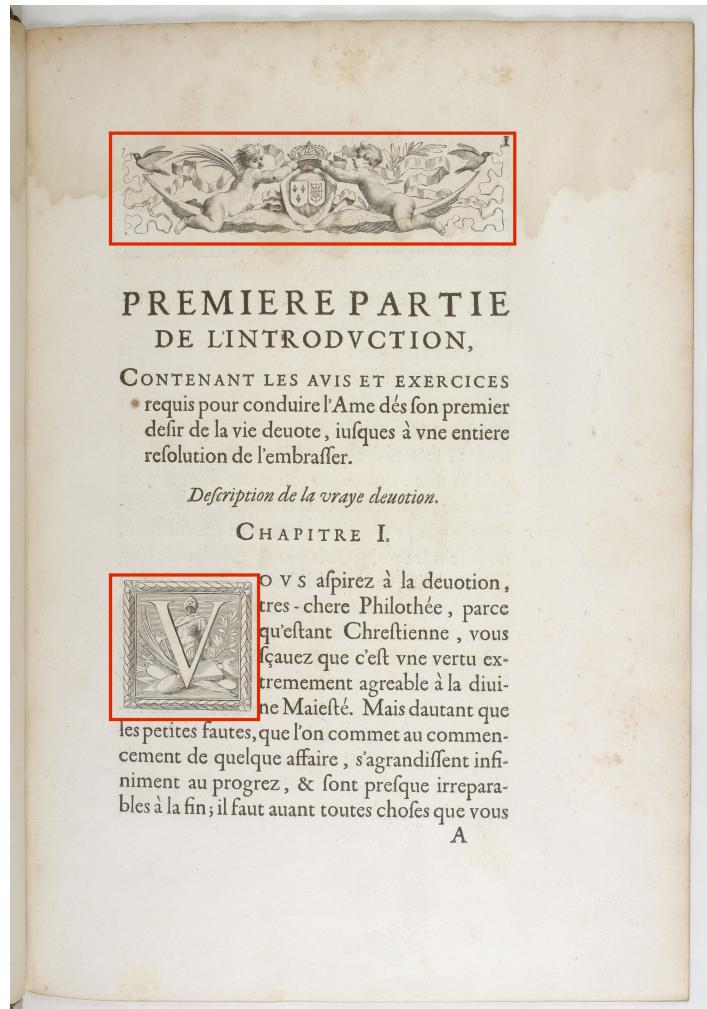


Segmentation: lignes

87545 BIGOTTINI (Émilie), célèbre danseuse, qui eut de grands succès à l'Opéra. — L. a. s. à M^{me} Guyet; 29 août 1833, 2 p. in-folio. 10 » Chaleureuse recommandation en faveur d'une femme de chambre.

L'OCR fonctionnant au niveau des lignes, il est fondamental de les extraire au mieux. L'utilisation d'outil dédié à la segmentation, et non d'un segmenteur intégré à l'OCR, peut être intéressant

Segmentation: la décoration



Le segmenteur peut extraire plus que des lignes: il peut extraire, par exemple, des ornements (bandeaux, initiales, culs-de-lampe...)

Principes d'apprentissage machine

Apprendre pour une machine

Pour faire faire quelque chose à une machine, on a deux possibilités:

1. Donner des règles à la machine: *suivons* a pour lemme *suivre*. Le problème c'est qu'il existe des cas ambigus: *suis*? Il faut alors rajouter des règles toujours complexes
2. Donner des exemples à la machine, qui va déduire des règles à partir des exemples: *je suis un homme -> être vs je suis le cours -> suivre*

L'IA c'est quoi?

L'intelligence artificielle se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d'« apprendre » à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune.

Il existe plusieurs méthodes:

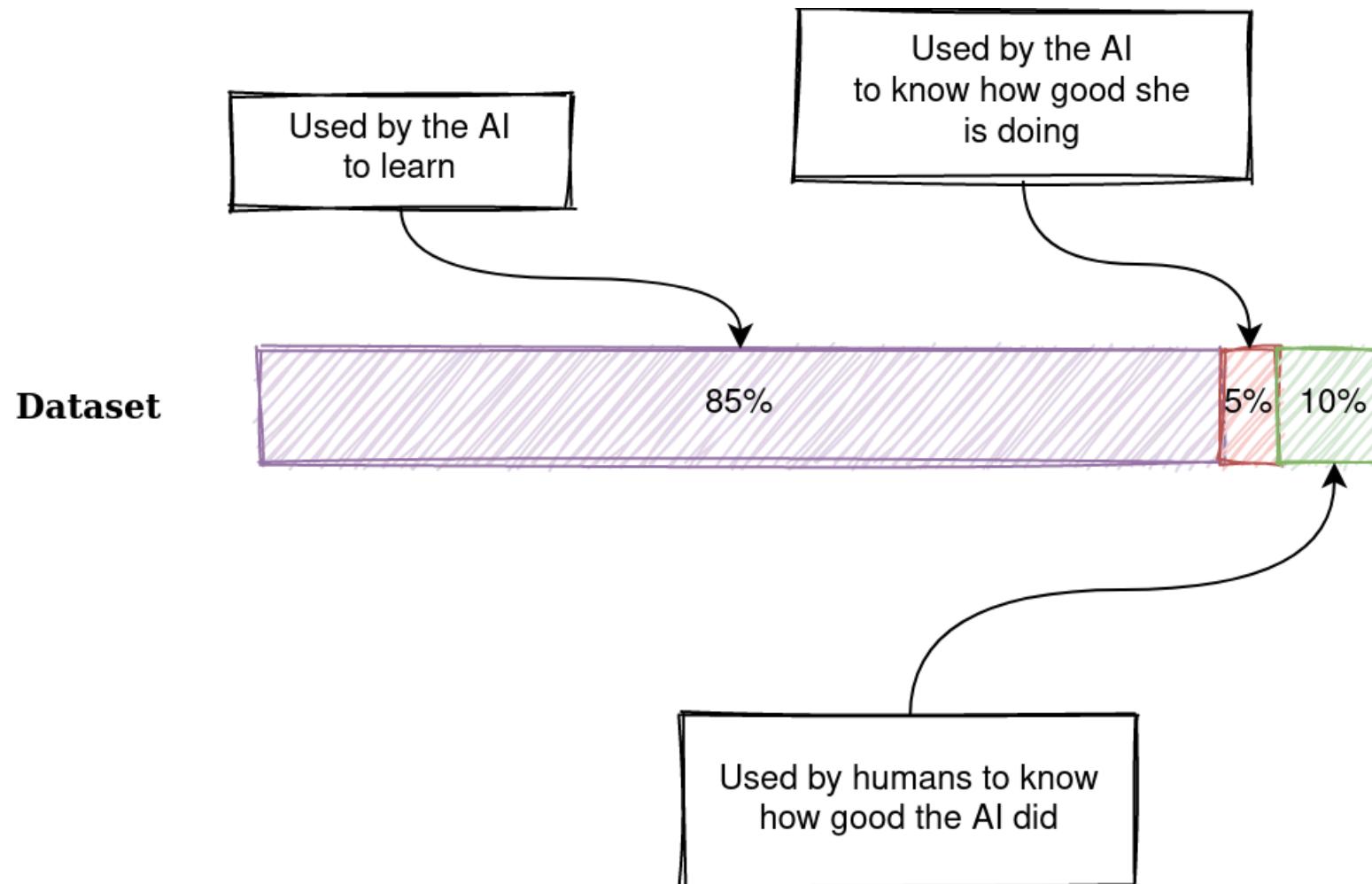
- Machines à vecteurs de support
- Méthodes statistiques
- Les réseaux de neurones
- ...

Préparer un dataset

Il va nous falloir trois jeux de données:

- Pour l'entraînement: les données sont regardées par la machine pour déduire les règles
- Pour la validation: les données sont utilisées pour contrôler l'apprentissage sur les données d'entraînement à chaque itération, avec le risque que la machine se mette à bachoter
- Pour le test final: les données, qui n'ont jamais été vues pendant l'entraînement, vont permettre de contrôler la qualité finale

Répartition du dataset



Le règne de la quantité

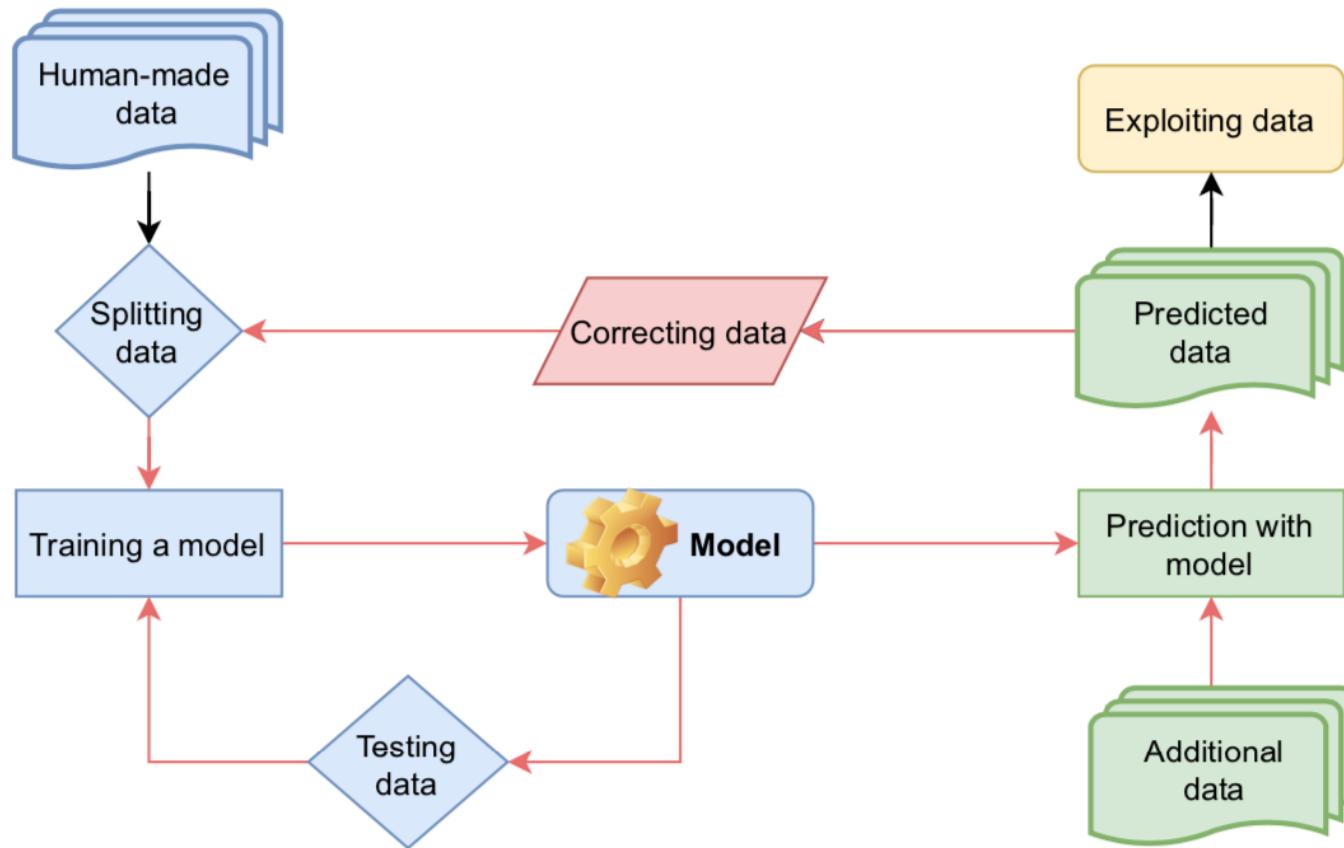
L'IA est extrêmement gourmande en données: le plus c'est le mieux, il en faut donc un maximum. Il peut donc être utile de mutualiser les données, comme le propose le projet *HTR United*:

<https://htr-united.github.io>

Mais pour partager des données, il faut qu'elles soient compatibles entre elles, tant du point de vue du format que de l'annotation:

- On ne peut pas mélanger des données en ALTO et en PAGExml
- Si certains transcrivent *v* et d'autres *u* c'est potentiellement un problème
- Si certains distinguent les colonnes et d'autre pas ça ne va pas
- ...

Roadmap pour l'entraînement



Petit lexique

- Modèle: somme de ce qui a été appris par la machine au cours de l'apprentissage
- Itération: cycle d'entraînement pendant lequel l'ordinateur voit toutes les données d'entraînement
- Surentraînement (*overfitting*): à force d'être évaluée sur le même jeu de validation à chaque itération, la machine se met à n'apprendre que pour passer ce test
- Généralisation: capacité du modèle à traiter des données qu'il n'a jamais vues.
- Arrêt prématué (*early stopping*): arrêt de l'entraînement avant le surentraînement
- *Fine tuning*: on repart depuis un modèle déjà entraîné, auquel on ajoute une couche d'information

Le résultat

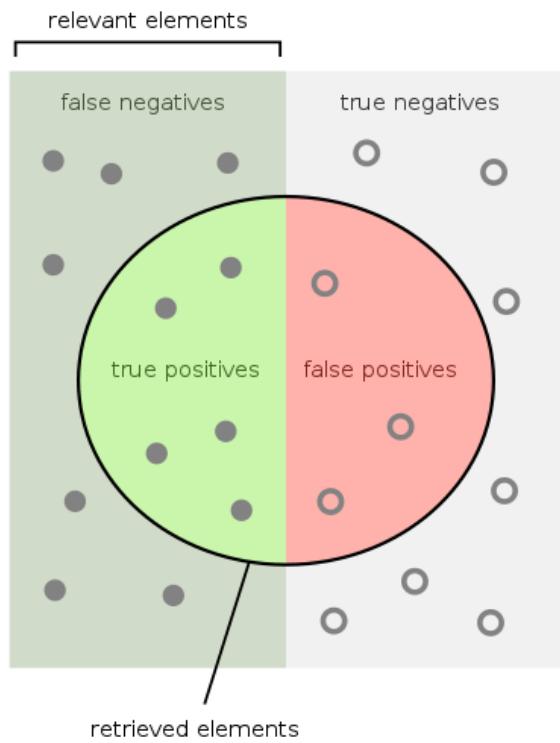
On utilise pour évaluer un modèle un score dit "F1", calculé à partir de la précision (*accuracy*) et du rappel (*recall*):

- Le rappel calcule le nombre de positifs bien prédit par notre modèle.
- La précision calcule le nombre de prédictions positives bien effectuées.

On parle donc:

- De faux négatif, quand le résultat est déclaré positif mais est en réalité négatif. La machine s'est trompée.
- De vrai négatif, quand le résultat est déclaré négatif mais est en réalité positif. La machine a oublié quelque chose.

Score F1



How many retrieved items are relevant?

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

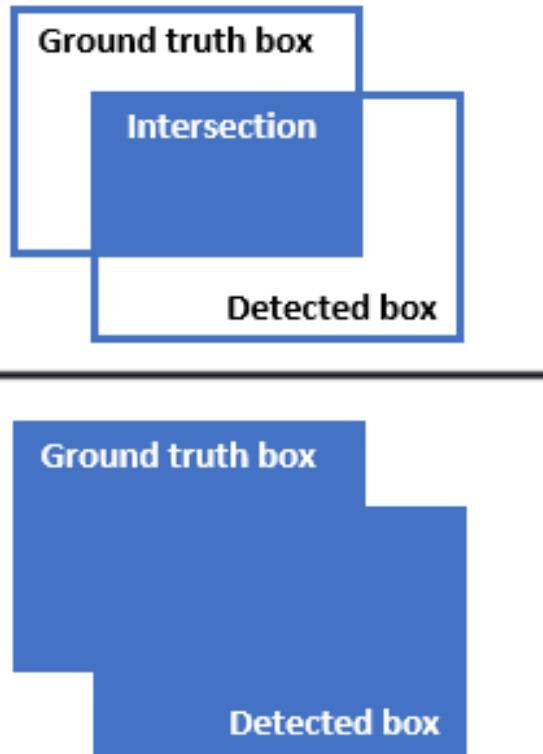
How many relevant items are retrieved?

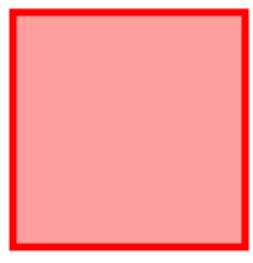
$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

IoU

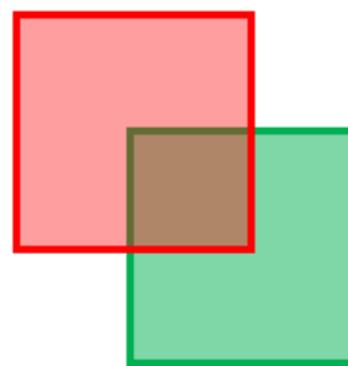
On utilise aussi l'intersection sur l'union (*Intersection over union*), plus adapté pour des zones (sur une page par exemple).

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{\text{Intersection}}{\text{Union}}$$

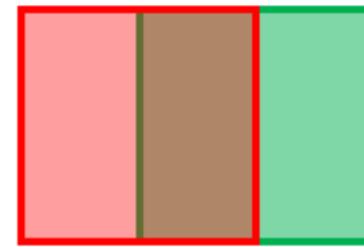




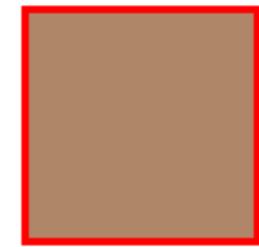
IoU = 0



IoU = 0.142



IoU = 0.333



IoU = 1

Sur le test

Le score du test n'a pas de valeur en soi, elle dépend:

- Des données d'entraînement
- Des données dans le jeu de test

99% sur 10 000 images ne vaut pas (forcément) 99% sur 1000 images

Pour le test, on peut utiliser des données:

- *In domain*: elles sont proches du jeu d'entraînement (tirées du même livre par exemple)
- *Out of domain*: elles sont différentes du jeu d'entraînement -- la question étant de savoir jusqu'à quel point différentes

Exercice

Regardez le dossier **Exercices** et tenter de trouver un dénominateur commun à toutes ces images

SegmOnto

Intuition



128 **Les Caractères**

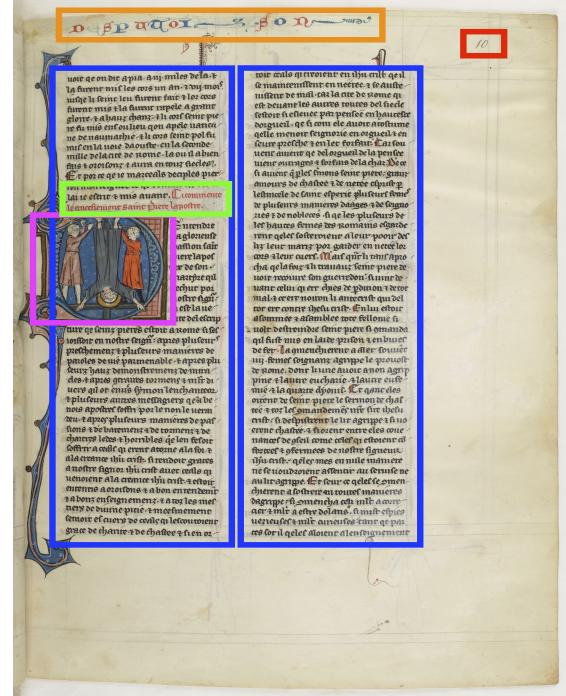
Dieux ; esperez donc toutes choses heureuses de cette Deesse : Après avoir parlé ainsi , il se retire dans sa maison où il fait un long récit à sa femme de la manière dont toutes choses se sont passées , & comme elles luy ont réussi au delà de ses souhaits.

D E L'AVARICE.

Ce vice est dans l'homme un oubli de l'honneur & de la gloire , quand il s'agit d'éviter la moindre dépense . Si un homme a remporté le prix de la* tragédie , il consacre à Bacchus des guirlandes ou des banderoles faites avec de l'écorce de bois , & il fait graver son nom sur un présent si magnifique . Quelquefois dans les temps difficiles le peuple est obligé

* qu'il a faite ou reçue.

Principes



Titre courant, numéro de page, corps du texte, initiale, rubrique...

Approche générale plutôt que spécifique

Mise en page plutôt qu'analyse sémantique

Une standardisation est cruciale pour la numérisation massive des documents qui s'annonce

Double objectif

- Partager les données (*upstream*): l'intelligence artificielle a besoin de données: les équipes de recherche ont besoin de partager des documents annotés pour améliorer les résultats de l'HTR en augmentant la quantité de données d'entraînement
- Partager les données (*downstream*): les équipes de recherche ont aussi besoin de mettre en commun des post-traitement de données pour l'exploration des corpus et la production/transformation automatisée de documents (TEI, RDF...)

→ Trouver un vocabulaire partagé et contrôlé pour décrire la page

Trouver un vocabulaire

Plutôt de créer un vocabulaire, il est préférable de s'appuyer sur de l'existant

- Vocabulaire PAGE xml, avec très peu de catégories
- Vocabulaire *Codicologia*, un vocabulaire multilingue pour la description des manuscrits, avec énormément de catégories, et adapté pour le RDF (Vocabulaire international de la codicologie - SKOS)

Liens:

- <http://codicologia.irht.cnrs.fr/>
- <http://gams.uni-graz.at/archive/objects/o:voccod/methods/sdef:SKOS/get>

Zones

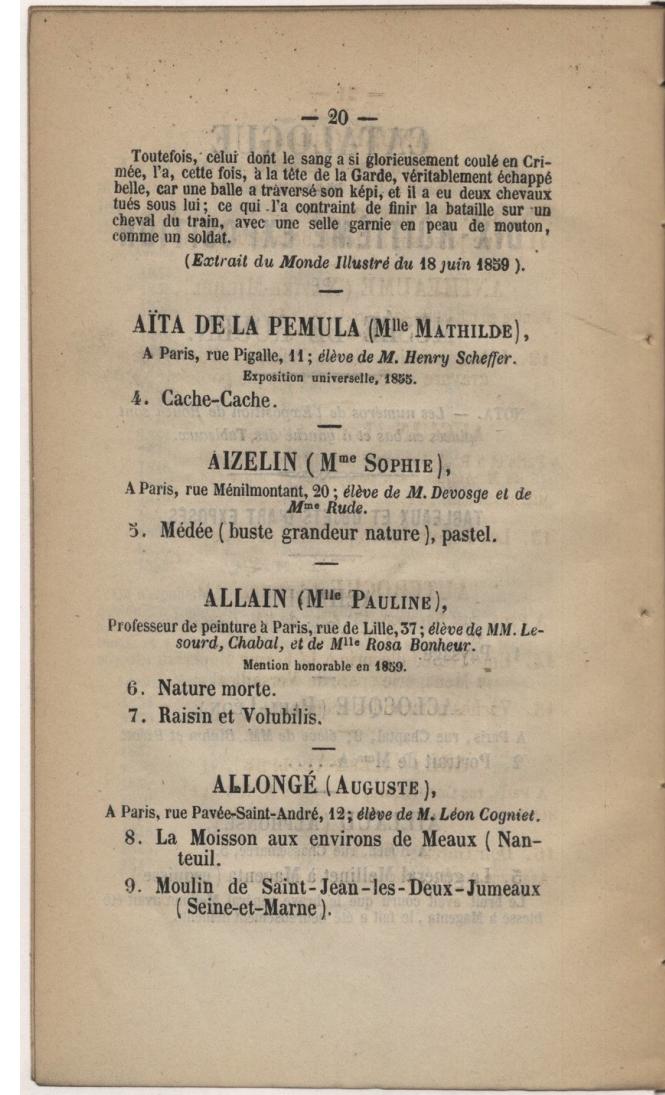
Zones	Zones (suite)
CustomZone	NumberingZone
DamageZone	QuireMarksZone
DigitizationArtifactZone	RunningTitleZone
DropCapitalZone	SealZone
GraphicZone	StampZone
MainZone	TableZone
MarginTextZone	TitlePageZone
MusicZone	

Lignes

- CustomLine
- DefaultLine
- DropCapitalLine
- HeadingLine
- InterlinearLine
- MusicLine

CustomZone

- Definition: characterises any kind of zone not fitting in the other categories, according to any convenient typology the user chooses.
- Subtypes: Any
- Examples: encoding catalogue entries with
 - CustomZone:entry#1
 - CustomZone:entry#2



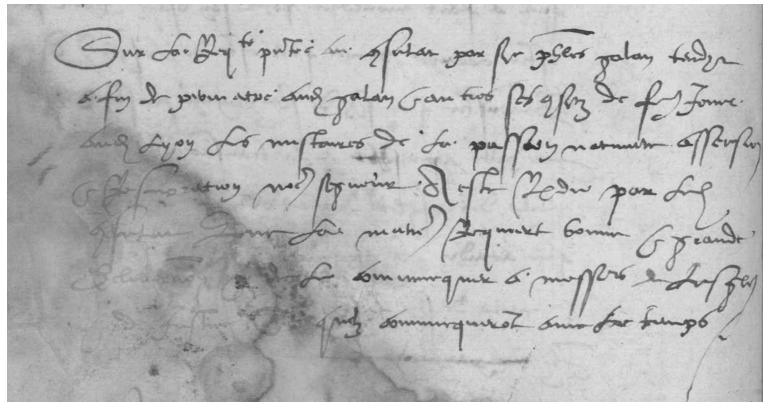
DamageZone

- Definition: characterises any area containing damage to the source, such as holes in the material (parchment, paper...), blots, etc.
 - Subtypes: suggested values include:
 - DamageZone:corrosion (*corrosion*)
 - DamageZone:hole (*trou*)
 - DamageZone:mold (*moisissure*)
 - DamageZone:peeled (*desquamé*)
 - DamageZone:soaked (*détrempé*)
 - DamageZone:scuffed (*eraillé*)
- ...

DamageZone 2

Examples:

DamageZone : soaked



DamageZone : hole



Identifying damaged area might prove useful, as they can affect the result of text prediction.

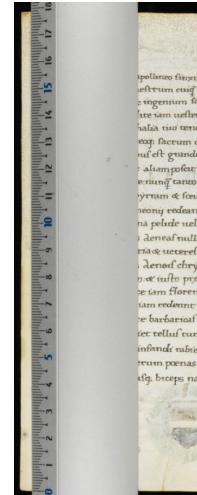
DigitizationArtefactZone

- Definition: it contains any type of item external to the document itself, but due to the process of digitisation, such as rulers or color tables added to help analyse the image.
- Subtypes: suggested values include:

DigitizationArtefactZone:testCard



DigitizationArtefactZone:ruler

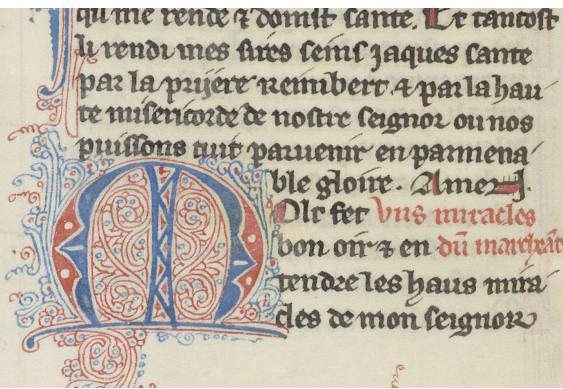


DropCapitalZone

- Definition: contains any type of initial letter, occupying a space corresponding to several lines of the main text or bearing significant ornamentation, be they historiated, ornated, flourished or painted initials (and excluding the following text line).
- Subtypes: suggested values include:
 - DropCapitalZone:historiated
 - DropCapitalZone:floriate
 - DropCapitalZone:flourish
 - DropCapitalZone:voided
 - DropCapitalZone:parted

DropCapitalZone

- DropCapitalZone:champ
- DropCapitalZone:facetted
- DropCapitalZone:plain
- DropCapitalZone:interlocked

Initiale manuscrite	initiale imprimée
 <p>qu me rende & domit cante. It tancost le rendi mes fures sens jaques cante par la priere rembert & par la hau- te misericorde de nostre seignor ou nos puissons qui paruenir en parmena- ble gloire. Ame. Olc fet vns miraclos bon onz en dñ marchat rendre les haus miraclos de mon seignor</p>	<p>LA sotte envie de discourir vient d'une habitude qu'on a contractée de parler beaucoup.</p>

GraphicZone

- Definition: characterises a zone containing any type of graphic element, from purely ornamental to consubstantial to the text (e.g., full page paintings, line-fillers, marginal drawings, figures, etc.).
- Subtypes: suggested values include:
 - GraphicZone:illustration
 - GraphicZone:ornamentation
 - GraphicZone:figure

GraphicZone

GraphicZone: illustration

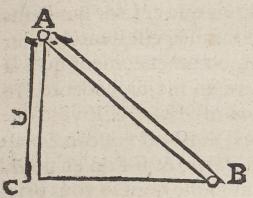


GraphicZone: ornamentation

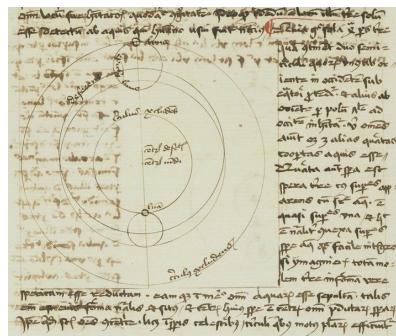


GraphicZone: figure

pend de sa pente & de sa longueur : Comme par le Tuyau A B, elle ira de mesme vitesse que par le Tuyau A C. Et pour fçauoir de quelle vitesse elle ira en cettuy- cy, il faut penser que la goutte d'Eau qui est vers C, a inclination à descendre aussi vite que si elle auoit desia descendu en l'Air libre depuis A iusques à C, & que la goutte qui est vers D, n'a inclination à descendre que de la mesme vitesse qu'elle auoit acquise en descendant en l'Air libre depuis A iusques à D , & ainsi des au-

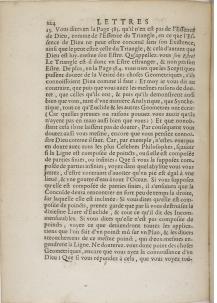


GraphicZone: figure



MainZone

- Definition the main area (text column) designed to contain text, either as a single or several columns (as designed in the conception of the layout: including eventually text, music notations, illumination, etc.).
- Subtypes: suggested values include:

MainZone:column	MainZone:column#1 and MainZone:column#2
	

MarginTextZone

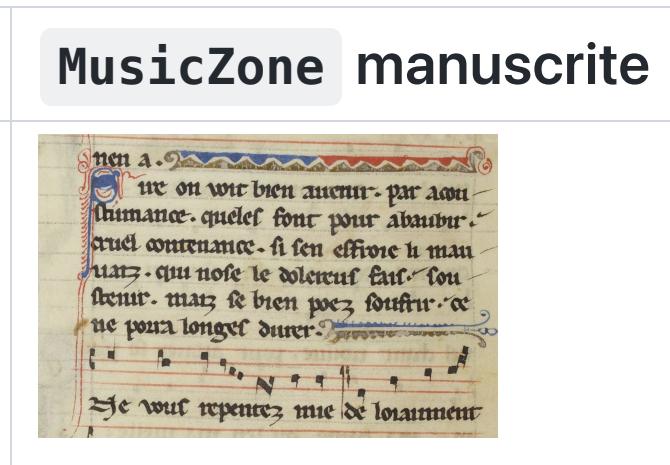
- Definition: characterises any **text zone** contained in the margins (upper, lower, inner or outer), including the space between two columns, whatever their semantic status (gloss, additions, ...).
- Subtypes: suggested values include:
 - MarginTextZone:note
 - MarginTextZone:commentary
 - MarginTextZone:correction
 - MarginTextZone:variants

MarginTextZone

MarginTextZone : variantes	MarginTextZone : variantes
<p>70 GOSOKI HAZAKENO</p> <p>باترولوج ٢٧٣ طرفهات معك هذه الطفهات معك مبتدا الخطات منك غزيره من الحمد ويلوين هذه الطفهات معك وتوّفي ٤٠ غزيره لجهة الحشد لا ياخذ من الماء والثلج عافر ودم العرس وبلغه في هنا على واكين لغرس مرتنا وانا يكون حملنا ٦٨١ ايسن كافتكم الكنية من اكين الايام ٦٩٣ يكون الفرج من اسلطفها فنا تلبية بشيئ الععن دعم العدن</p>	<p>باترولوج ٢٧٣ طرفهات معك هذه الطفهات معك مبتدا الخطات منك غزيره من الحمد ويلوين هذه الطفهات معك وتوّفي ٤٠ غزيره لجهة الحشد لا ياخذ من الماء والثلج عافر ودم العرس وبلغه في هنا على واكين لغرس مرتنا وانا يكون حملنا ٦٨١ ايسن كافتكم الكنية من اكين الايام ٦٩٣ يكون الفرج من اسلطفها فنا تلبية بشيئ الععن دعم العدن</p>

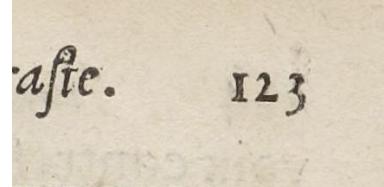
MusicZone

- Definition: characterises an area containing musical notations, such as neumes, staves, etc., with the possible inclusions of text.
- Subtypes: none
- Examples:

MusicZone imprimée	MusicZone manuscrite
 <p>BELLEROPHON TRAGEDIE ACTE PREMIER. Le Théâtre Représente la Ville de Patare Capitale du Royaume de Lycie. Scene Première. Senobé. Stanobé. Arrie. Non, les Soulouemens d'une Ville re... belle ne</p>	 <p>nen a. Ne on wit bien auenur. par aon stumance. quelles font pour abanir. cruel contenance. si sen effroie n mai juaz. qui nose le dolereus fai. sou auenur. matz le bien poez souffrir. ce ne poura longes durer. Je vous repentez mie de lorainment</p>

NumberingZone

- Definition: it characterises a zone containing the page number.
- Subtypes: suggested values include:
 - NumberingZone:page
 - NumberingZone:other
- Examples

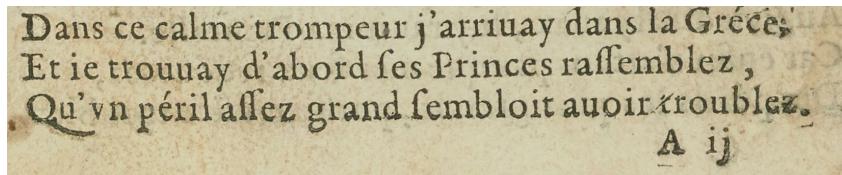
NumberingZone:folio	NumberingZone:page
	

QuireMarksZone

- Definition: characterises a zone containing a quire signature (i.e., a *ii*), catchword, or any kind of element relative to the material organisation of the source, with the exclusion of page numbers.
- Subtypes: suggested values include:
 - QuireMarksZone:signature
 - QuireMarksZone:catchwords

QuireMarksZone

- Examples
 - QuireMarksZone:signature

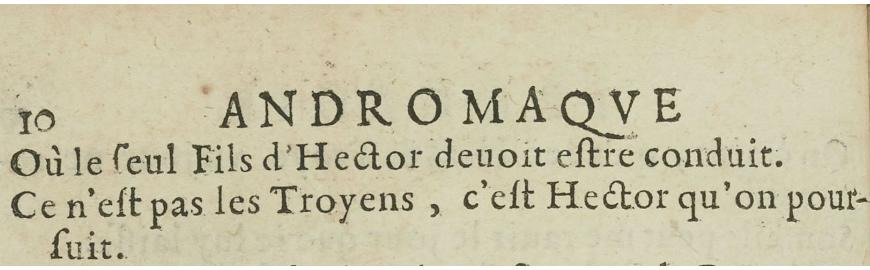


- QuireMarksZone:catchwords

QuireMarksZone manuscrite	QuireMarksZone imprimée
<p>Chere Cleonne cours.</p>	<p>Chere Cleonne cours. Ma vangeance est perduë, S'il ignore, en mourant, que c'est moy qui le tuë.</p> <p>CLEONNE.</p>

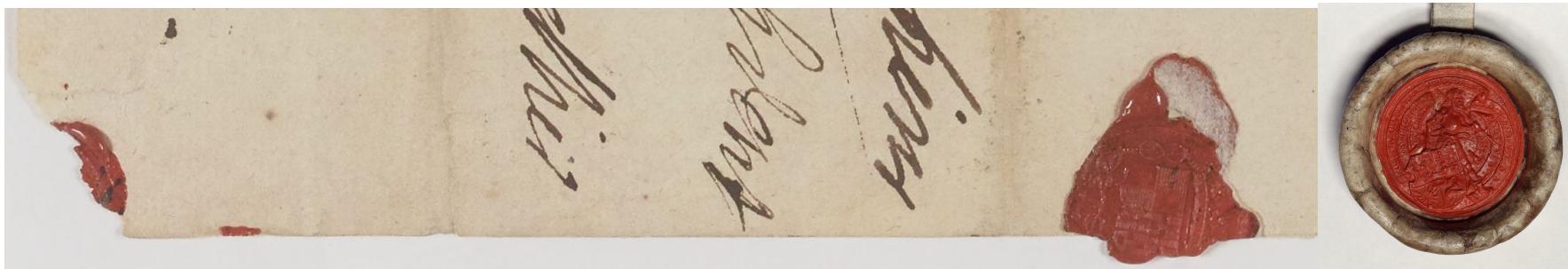
RunningTitleZone

- Definition: characterises a zone containing a running title.
- Subtypes: none
- Examples

RunningTitleZone imprimée	RunningTitleZone manuscrite
 <p>10 ANDROMAQUE Où le seul Fils d'Hector deuoit estre conduit. Ce n'est pas les Troyens , c'est Hector qu'on poursuit.</p>	 <p>...od p[er]d[er]it tous adieu Il faisoit aumosnes il[le]t noit son cors en chaete & li gardoir si la osience purielement de la ou il estoit en ...[red ink] a[nd] p[er]oures nostre seignor resucita[re] Et or pri e nostre signor qui most de la cheualerie de cest siecle & l'ame doinst le vntel cheualier</p>

SealZone

- Definition: it characterises a zone containing a seal.
- Subtypes: none
- Examples

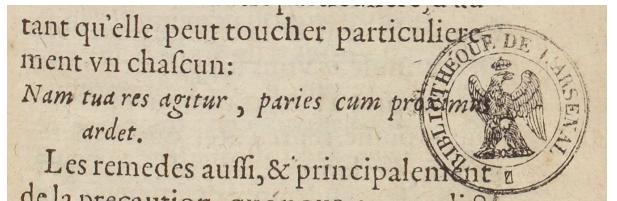


StampZone

- Definition: it characterises a zone containing a stamp, be it a library stamp or a mark from a postal service.
- Subtypes: suggested values include:
 - StampZone:postal
 - StampZone:curatorial

StampZone

- Examples
 - StampZone: curatorial



- StampZone: postal



TableZone

- Definition: characterises a zone containing a table of any kind
 - Subtypes: none
 - Examples

TitlePageZone

- Definition: it characterises a zone containing a title distinct from the main text. It is mainly used for prints.
- Subtypes: none
- Examples

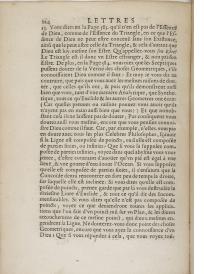
TitlePageZone manuscrite	TitlePageZone imprimée
	 <p>ADVIS SUR LA PESTE.</p> <p>A PARIS, Chez DAVID DOYCEY, Libraire au Rond Point du Mercuriale. M. DCVII.</p> <p>AVEC PERMISSION.</p>

CustomLine

- Definition: characterises any kind of line not fitting in the other categories, according to any convenient typology the user chooses.
- Subtypes: Any
- Justification: all projects have specific needs regarding types of lines peculiar to their sources or their goals and not covered by the standard types.

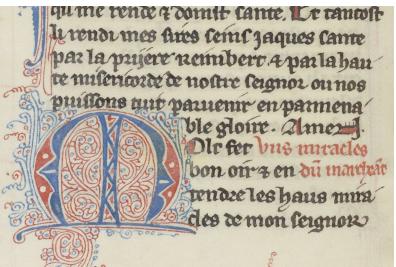
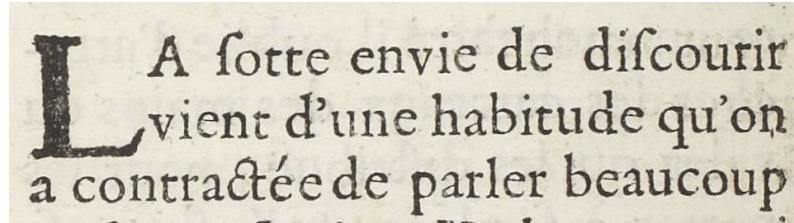
DefaultLine

- D*efinition: it characterises any kind of standard text line, whether they are included in the `MainZone` text, in the `MarginZone`, in `MusicZone`, or in any type of zone.
- Subtypes: none
- Examples

DefaultLine imprimée	DefaultLine manuscrite
	

DropCapitalLine

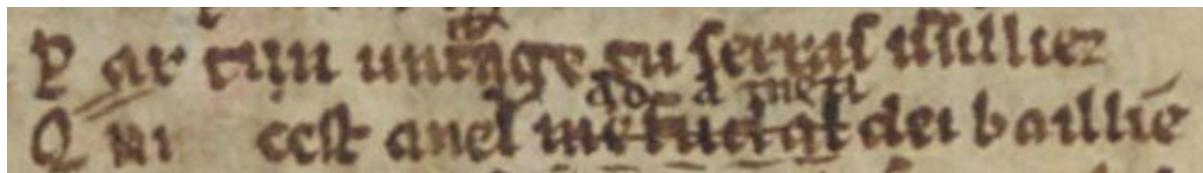
- Definition: characterises a line on which rests a **DropCapital**.
- Subtypes: none
- Examples

DefaultLine imprimée	DefaultLine manuscrite
 A printed page showing a drop capital 'D' at the beginning of a line of text. The text is in a black serif font, and the drop capital is a large, ornate letter.	 A manuscript page showing a drop capital 'L' at the beginning of a line of text. The text is in a black Gothic script, and the drop capital is a large, decorated letter.

- Justification: drop capitals are both specific zones, and bear a text.

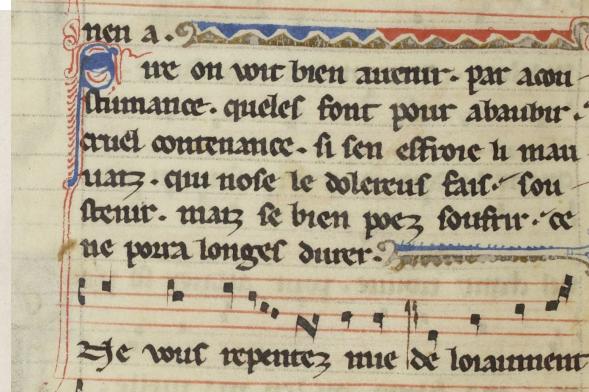
InterlinearLine

- Definition: characterises a line that is not a standard text line, but has been added between two of them, for instance to include a forgotten word.
- Subtypes: suggested values include:
 - InterlinearLine:commentary
 - InterlinearLine:correction
- Examples
 - InterlinearLine:correction



MusicLine

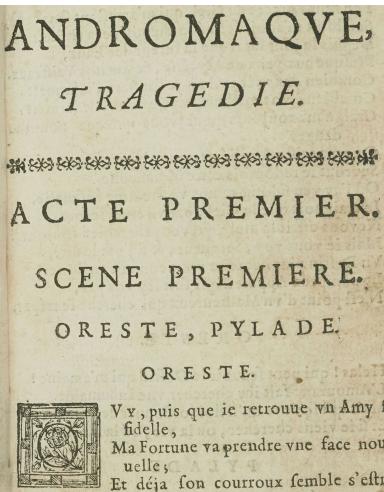
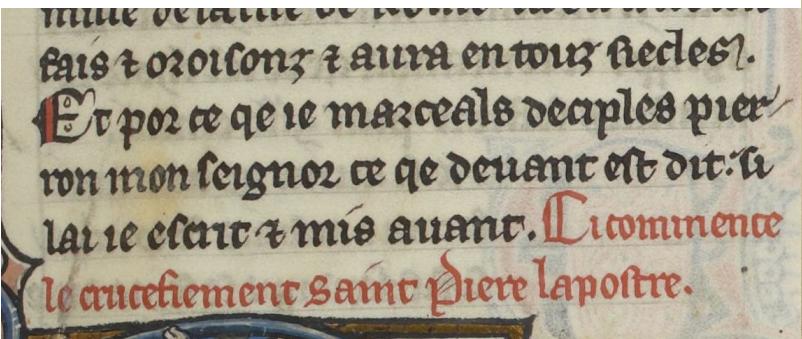
- Definition: characterises the central line of a musical stave.
- Subtypes: none
- Examples



- Justification: analysis or prediction of musical content can be as necessary as the text's.

HeadingLine

- Definition: characterises a line containing a rubric, for instance signalling the beginning of a new text.
- Subtypes: suggested values include:
 - HeadingLine:incipit
 - HeadingLine:explicit
 - HeadingLine:title
- Examples



Exercice : créer un nouveau document

La description du document (1)

Après avoir cliqué sur `Create new Document`, plusieurs éléments sont à préciser dans la partie *Description* :

- commencer par donner un nom à votre document ;
- *Main script* : Latin ;
- *Read direction* : *Left to Right* ;
- *Line offset* : *Baseline*
- *Metadata* : les métadonnées sont remplies automatiquement si vous choisissez de charger votre document via IIIF.

La description du document

← Description Images Edit Models Document created successfully! X

Contrat_de_mariage_Parlement_Paris

Project Mazarinades

Main script Latin

Read direction Left to right

The read direction describes the order of the elements in the document, in opposition with the text direction which describes the order of the words in a line and is set by the script.

Line offset Baseline

The position of the line relative to the polygon.

Metadata

Key	Value
-----	-------

Ontology ?

The default configuration only contains semantic types. X

Métadonnées remplies automatiquement via IIIF

Metadata

Contenu numerisé	Intégral	x
doc_id	Mazarinades OBVIL lot2	x
Catégorie	[{@value': 'Livres imprimés'}, {@value': 'Mazarinades'}]	x
Type de document	[{@value': 'text'}, {@value': 'language materials'}]	x
Droits	[{@value': 'Voir ']	x
Notice complète	[{@value': '']	x
Référence	[{@value': 'M 15433-46'}, {@value': '751063002_MAZ015433_000046']}	x
Cote	Bibliothèque Mazarine, M 15433-46	x
Description matériel	[{@value': 'Format : in-4'}, {@value': '16 p.'}, {@value': 'Support\xA0: p']	x
Nombre de vues	41	x
Titre	Contract de mariage du Parlement, avec la ville de Paris.	x
Autre titre	[{@value': '[Mazarinade. 1649]'}, {@value': 'Contrat de mariage du Par']	x
Sujet	[{@value': 'France. Parlement de Paris'}, {@value': 'Paris (France) -- 16']	x
Langue	fre	x
Lieu de fabrication ci	M. DC. XLIX.	x
Ville	Orléans	x
Pays	France	x
Coordonnées GPS	47.902964,1.909251	x
Date	[{@value': 'M. DC. XLIX.'}, {@value': '1649'}]	x

La description du document (2)

La dernière étape consiste à mettre à jour l'ontologie du document afin qu'elle corresponde à celle des deux modèles précédemment chargés :

Region types :

- ajouter manuellement *CustomZone*, *DamageZone*, *DigitizationArtifactZone*, *DropCapitalZone*, *GraphicZone*, *MainZone*, *MarginTextZone*, *MusicZone*, *NumberingZone*, *QuireMarksZone*, *SealZone*, *StampZone*, *TableZone* et *TitlePageZone* (15 éléments au total).

Line types :

- ajouter manuellement *CustomLine*, *DefaultLine*, *DropCapitalLine*, *InterlinearLine*, *MusicLine* and *HeadingLine* (6 éléments au total).
- Enfin, appuyer sur `Create`.

L'ontologie

Ontology



The default configuration only contains semantic types.



It is not recommended to mix semantic, stylistic and layout types.

It is required to hit 'Update' to add your new types to the document.

Region types

- Commentary Illustration Main Title CustomZone DamageZone
- DigitizationArtefactZone DropCapitalZone GraphicZone MainZone
- MarginTextZone MusicZone NumberingZone QuireMarksZone SealZone
- StampZone TableZone TitlePageZone



Line types

- Correction Main Numbering Signature CustomLine DefaultLine
- DropCapitalLine InterlinearLine MusicLine RubricLine



La segmentation des images

Images

eS eScriptorium Home Contact My Projects My Models My Cluster Jobs Hello abartz ▾

← Description Images Edit Models Contrat_de_mariage_Parlement_Paris

Drop images here or click to upload.

Select all Unselect all Selected 0/38 Import Export Train Binarize Segment Transcribe

1 2 3 4 5 6 7 8 9

1: A dark, textured page with a small white mark at the bottom right.
2: A page featuring a large, ornate coat of arms in red and gold.
3: A page with faint, illegible handwritten text.
4: A page with faint, illegible handwritten text.
5: A page with printed text: "DE MARIAGE DV PARLEMENT, AVEC LA VILLE DE PARIS".
6: A page with printed text: "CONTRACT DE MARIAGE".
7: A page with printed text: "LE RÉGNE DE FRANCE".
8: A page with printed text: "LE RÉGNE DE FRANCE".
9: A page with printed text: "LE RÉGNE DE FRANCE".

Regions

- Commencer par sélectionner l'ensemble des images que vous souhaitez segmenter ;
- Cliquer sur Segment ;
- Le Model à utiliser est le modèle de segmentation par défaut (`blla.mlmodel`) ;
- Sélectionner Regions dans Segmentation steps ;
- Lancer la segmentation et patienter.

Paramètres pour la segmentation par régions

20 image(s) selected.



Beware that re-segmenting will delete any existing transcriptions!



Select a model

Model

default (blla.mlmodel)



Segmentation steps

Regions



Text direction

Horizontal l2r



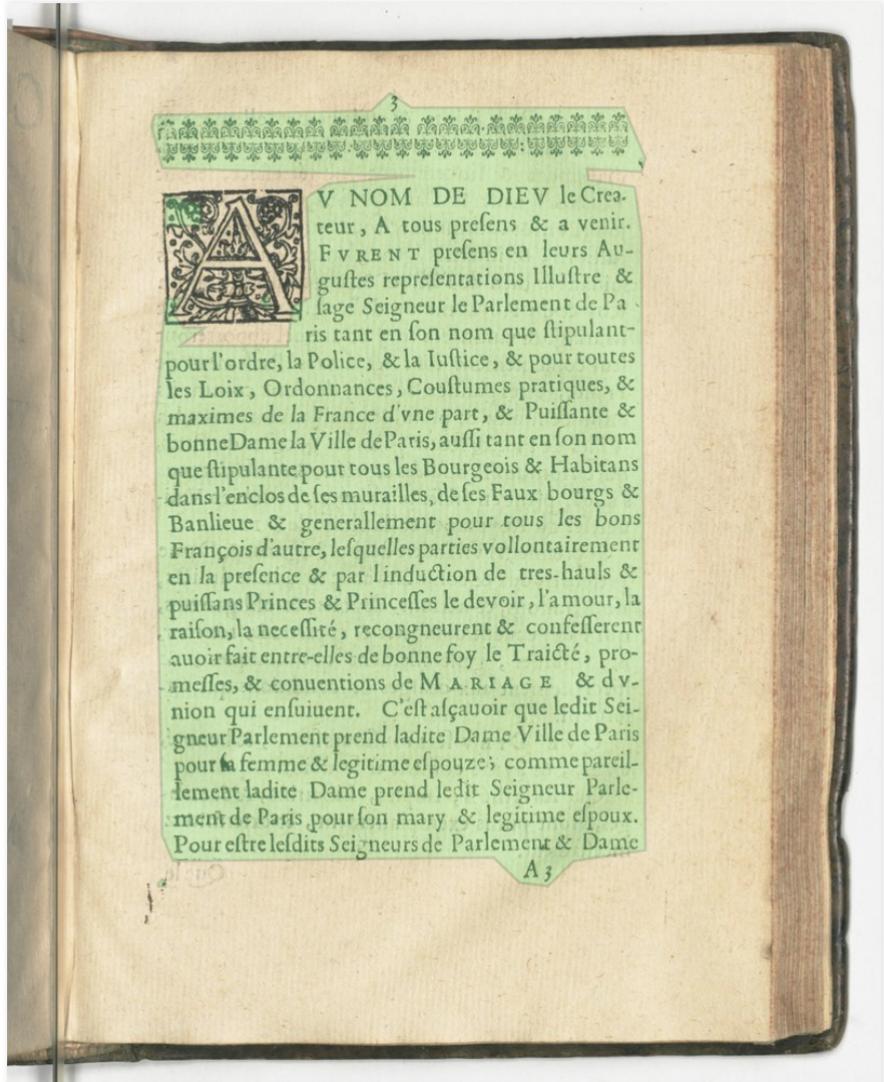
Override

If checked, deletes existing segmentation **and bound transcriptions first!**

Close

Segment

Exemple de résultat



Corrections

- Vous pouvez à présent consulter les résultats de la segmentation en cliquant sur **Edit**.
- Le modèle par défaut n'étant pas familier de l'ontologie que nous souhaitons utiliser, il reste à associer chacune des zones repérées à la bonne terminologie.
- Il se peut qu'il soit nécessaire de redéfinir certaines zones, pour cela cliquer sur le petit ciseau pour passer en mode *édition*.

Lines

Il reste à présent à repérer les lignes du document :

- sélectionner les images ;
- cliquer sur Segment ;
- le Model à utiliser est le modèle par défaut ;
- sélectionner Lines Baselines and Masks dans Segmentation steps ;
- lancer la segmentation ;
- vérifier que les lignes ont bien été reconnues.

N.B. : il est important de commencer par segmenter les pages par régions puis par lignes afin que ces dernières soient automatiquement associées à leurs zones.

Corrections

- Vous pouvez à présent consulter les résultats de la segmentation en cliquant sur **Edit**.
- Le modèle par défaut n'étant pas familier de l'ontologie que nous souhaitons utiliser, il reste à associer chacune des lignes repérées à la bonne terminologie.
- Il se peut qu'il soit nécessaire de retracer certaines lignes, pour cela cliquer sur le petit ciseau pour passer en mode *édition*.